

# Degree Correlations in Random Geometric Graphs

A. Antonioni\* and M. Tomassini†

*Information Systems Department,  
Faculty of Business and Economics,  
University of Lausanne, Switzerland*  
(Received: date / Accepted: date)

Spatially embedded networks are important in several disciplines. The prototypical spatial network we assume is the Random Geometric Graph of which many properties are known. Here we present new results for the two-point degree correlation function in terms of the clustering coefficient of the graphs for two-dimensional space in particular, with extensions to arbitrary finite dimension.

PACS numbers: 02.10.Ox, 89.75.-k

In the last decade, thanks to abundant data, new models, and adequate software tools, complex networks have been thoroughly investigated in many disciplines and as a substrate of many phenomena. A synthesis is now emerging, as can be seen e.g. in the recent comprehensive treatment by Newman [1]. Most of this work has dealt with “relational” networks, i.e. graphs in which distances do not have physical meaning and are just dimensionless quantities measured in terms of edge hops. Indeed, many networks are mainly of this kind such as big social networks. However, in many cases the physical space in which networks are embedded and the actual distances between nodes are important such as in rail and road networks, ad hoc communication device networks, and other geographical and transportation networks. The recent comprehensive review by Barthélemy [2] has at last put together a large amount of scattered material on spatial networks. The Random Geometric Graph (RGG) is a standard spatial network model that plays a role for spatial networks similar to the one played by the Erdős-Rényi random graph for relational ones. This model is well known [2–4] but some of its second-order features have not yet been uncovered. Among these, there is the question of the degree correlation functions. In this Communication we present some results on degree correlations on RGGs that we believe were previously unknown.

The construction process of a RGG with  $N$  nodes and radius  $R$  can be summarized as follows [3, 4]:

- the  $N$  nodes are placed on the unitary space  $\Omega \in \mathbb{R}^d$  with uniform distribution.
- an edge is created for every pair of nodes whose distance is  $r < R$ .

In particular, the distance is given by some metric on  $\Omega$ .

In this work we have dealt with 2-dimensional RGGs and the Euclidean metric distance on  $\mathbb{R}^2$ . The unitary space  $\Omega$  is the square  $[0, 1]^2$  with no boundary conditions (torus). We call the *neighborhood area*  $V_X$  of node  $X$  the set of the points which are at distance lower than the radius  $R$  from  $X$ . In Fig. 1a three nodes  $X, Y, Z$  and their neighborhood

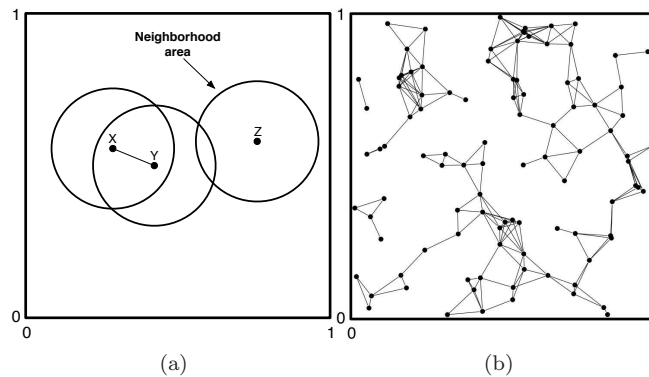


FIG. 1: (a) Neighborhood area. (b) An example of RGG with  $N = 100$  and  $R = 0.13$  (with boundary conditions).

\*alberto.antonioni@unil.ch

†marco.tomassini@unil.ch

areas have been depicted.  $X$  and  $Y$  are connected through an edge since they are within the neighborhood area of each other respectively, while  $Z$  is not connected to  $Y$  even if it is sharing a common area with it. Fig. 1b shows a RGG realization with  $N = 100$  and  $R = 0.13$ ; in this case, for illustrative purposes, we have considered the boundary conditions on  $\Omega$ .

The average degree  $\bar{k} = \bar{k}(N, R)$  of a RGG can be easily estimated by the formula  $\bar{k} = \rho V$ , where  $\rho$  is the node density, representing the number of nodes within a unit space, and  $V$  is the neighborhood area volume. In this case  $\rho = N$ , since  $\Omega$  is an unitary space, and  $V = \pi R^2$ . In conclusion,  $\bar{k} = \pi N R^2$ . According to this result, it is possible to consider  $\bar{k}$  as a parameter of RGGs, instead of the radius  $R$ . Therefore, in order to construct RGG with an average degree that tends to  $\bar{k}$  as  $N, 1/R \rightarrow \infty$ , it is sufficient to use the radius  $R = \sqrt{k/(\pi N)}$ .

The degree distribution of RGGs can be estimated regarding the probability density function of having a node  $X$  of degree  $k$ , given that there are other  $N$  nodes uniformly distributed in  $\Omega$ . More precisely,  $N - 1$  other nodes, but  $N \sim N - 1$  for large values of  $N$ . This probability follows the binomial distribution and it is equal to:

$$\mathbb{P}(k_X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

where  $p = \pi R^2$ , since it represents the proportion between  $V$  and  $\Omega$ . The Poisson distribution with parameter  $\lambda = Np$  can be used as an approximation of the binomial distribution if  $N$  is sufficiently large and  $p$  is sufficiently small. In this case the degree distribution will be approximated by:

$$\mathbb{P}(k_X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where  $\lambda = \bar{k}$ .

Fig. 2 shows three RGG cumulative degree distributions with the same number of nodes and different average degree.

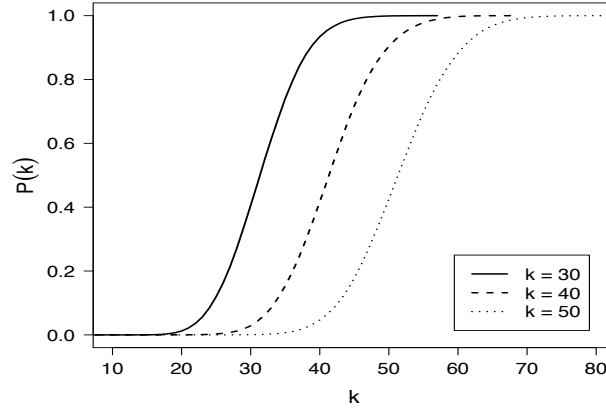


FIG. 2: Cumulative degree distribution functions for three realizations of RGG with  $N = 10000$  and three different average degrees.

The average clustering coefficient is given averaging on all node's individual clustering coefficients [1, 5]. This property on RGGs was extensively studied in the work of Dall and Christensen [4], in which they have found the law for the average clustering coefficient as a function of the dimension of the space. Here the dimension is equal to 2, and it is possible to demonstrate that the average clustering coefficient  $c_2$  tends to  $A_2 = 1 - \frac{3\sqrt{3}}{4\pi} \sim 0.5865$ , for large values of  $N$  and for all 2-dimensional RGGs [4] in the Euclidean space. By analogy, we shall call  $c_d$  the average clustering coefficient of a  $d$ -dimensional RGG.

This important result depends on the particular construction of RGGs. The average clustering coefficient tends to the ratio of the average shared neighborhood area of two connected nodes and the whole neighborhood area. It is clear that changing the radius  $R$  this fraction maintains the same value. This phenomenon, which conducts to a fixed average clustering coefficient for every RGGs, will be studied in depth in order to estimate the degree correlations in the following. Many other properties of RGGs have been studied in Penrose's book [3].

Due to its construction process, in a 2-dimensional RGG there is positive degree-degree correlation. This property is commonly detected studying the average degree of the neighborhood of a given node of degree  $k$  [6]. The function  $k_{nn}(k)$  (*nearest neighbor average degree*) represents the average degree of the neighborhood of all nodes of degree  $k$ . The properties which emerge from the spatial construction of RGGs allow us to evaluate  $k_{nn}(k)$  with a mean-field

method for very large values of  $N$ .

It is possible to evaluate the average degree of neighbors node estimating the average shared area of two connected nodes. Fig. 3 depicts the case of two connected nodes  $X$  and  $Y$ , and their shared area  $A$  (grey).  $B$  is the complementary area of  $A$  to neighborhood area  $V = \pi R^2$ , and  $r$  is the distance between the two nodes. The area  $C$  is symmetrical and equivalent to  $B$ .

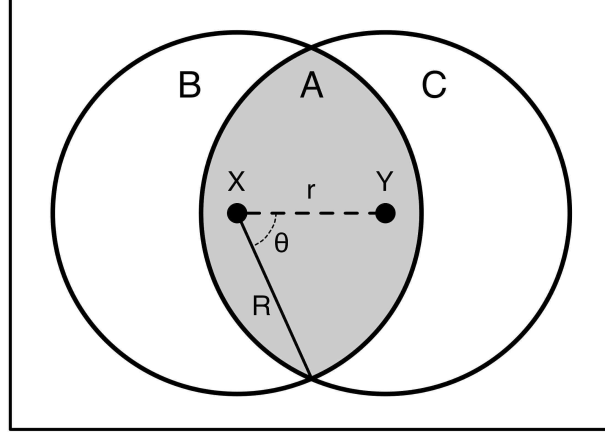


FIG. 3: Two connected nodes  $X, Y$  and their neighborhood areas of radius  $R$ . The distance between them is  $r$  and the angle is  $\theta = \arccos(\frac{r}{2R})$ . The grey area  $A$  is the shared area of the two connected nodes.  $B$  and  $C$  are the complementary areas of  $A$  to  $V$ .

The shared area  $A(r)$  of two connected nodes is only dependent on the distance  $r$  between them. The formula for  $A(r)$  can be simply derived from those of the circular sector and is equal to:

$$A(r) = R^2(2\theta - \sin 2\theta), \quad \text{where } \theta = \arccos\left(\frac{r}{2R}\right)$$

Thus, the average shared area  $\bar{A}$  is obtained integrating and averaging  $A(r)$  on all possible neighbors  $Y$  of  $X$  (Fig. 3).

We now calculate this area:

$$\bar{A} = \bar{A}(R) = \frac{\int_0^{2\pi} \int_0^R R^2(2\theta - \sin 2\theta) r dr d\phi}{\pi R^2}$$

where variables  $0 \leq r \leq R$  and  $0 \leq \phi \leq 2\pi$  represent all possible neighbors  $Y$  in the neighborhood area of  $X$ .

The integral in the numerator is calculated by using the substitutions  $r = 2R \cos \theta$  and  $\theta = \psi/2$ , which leads to:

$$-R^4 \int_0^{2\pi} \int_{\pi}^{\frac{2}{3}\pi} (\psi - \sin \psi) \sin \psi d\psi d\phi = R^4 \left( \pi^2 - \frac{3\sqrt{3}}{4\pi} \right)$$

It follows that the average shared area is:

$$\bar{A} = R^2 \left( \pi - \frac{3\sqrt{3}}{4} \right)$$

According to the last result, it is possible to evaluate the ratio of  $\bar{A}$  and  $V$ , which leads to:

$$\frac{\bar{A}}{V} = \frac{\bar{A}}{\pi R^2} = 1 - \frac{3\sqrt{3}}{4\pi} = A_2$$

where  $A_2 \sim 0.5865$  is the asymptotic value of the average clustering coefficient of 2-dimensional RGGs.

Finally, using the mean-field method to evaluate the neighbors shared area, it is possible to find the expression for the function  $k_{nn}(k)$ . In order to understand better this result, it is useful to use again the notation of Fig. 3. Focusing on node  $X$  of known degree  $k_X$ , we are trying to evaluate the average degree of its neighbor  $Y$ . This degree is given by two different areas, where the nodes density could be different.

The first area which brings neighbors to node  $Y$  is the shared area  $A$ , which is approximated by  $\bar{A}$  in the mean-field method, and where the nodes density is equal to  $\frac{k_X}{V}$ , thanks to the fact that we know the degree of  $X$ .

The second area in which node  $Y$  could find other neighbors is the complement  $C$  of  $A$ . In particular, approximating  $A$  by  $\bar{A}$  we approximate also  $C$  to  $\bar{C} = V - \bar{A}$ . In  $C$  the nodes density is equal to  $\rho = N$ , since we do not have more information about this area.

Putting together this information we can easily give the expression for  $k_{nn}(k_X)$ , which represents the degree of node  $Y$ :

$$k_{nn}(k_X) = \frac{k_X}{V} \bar{A} + \rho(V - \bar{A}) = A_2 \cdot k_X + (1 - A_2) \bar{k}$$

But  $k_X$  stands for the degree of a generic node in the graph and this leads to the final expression of the function:

$$k_{nn}(k) = A_2 \cdot k + (1 - A_2) \bar{k}$$

This function is linear in  $k$  and reveals the positive relation between the degree of a node and its average neighbors degree ( $A_2 > 0$ ).

The degree-degree correlation is not only estimated by  $k_{nn}(k)$ , but it is also exactly given by the *assortativity coefficient*, which is the Pearson correlation coefficient of degree between pairs of connected nodes. This value is widely used as a measure of the strength of linear dependence between two variables [1]. As for the average clustering coefficient, we use the notation  $r_2$  to indicate the assortativity coefficient of a 2-dimensional RGG.

As we have seen above, the function  $k_{nn}(k)$  is linear and, applying the mean-field method, it can represent the regression line for the two variables, degree ( $X$ ) and neighbor degree ( $Y$ ). We thus assume that  $k_{nn}(k)$  is the regression line and we derive  $r_2$  from that.

The regression line slope  $b$  tends to  $A_2$  for large values of  $N$  and is defined by the following formula:

$$b = \frac{\text{cov}(X, Y)}{\sigma_X^2} \rightarrow A_2, \quad N \rightarrow \infty$$

where  $\sigma_X$  is the standard deviation of variable  $X$  and  $\text{cov}(X, Y)$  is the covariance of the two variables  $X, Y$ .

On the other hand,  $r_2$  by definition is the covariance of the two variables  $X, Y$  divided by the product of their standard deviations:

$$r_2 = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

It follows this last result:

$$r_2 \rightarrow \frac{\sigma_X}{\sigma_Y} A_2$$

where  $\sigma_X, \sigma_Y$  are the standard deviations of variables  $X$  and  $Y$ , respectively.

In order to estimate these two standard deviations, we focus on their distribution functions  $f_X(x)$  and  $f_Y(y)$ . We already know that  $f_X(x)$  is a Poisson distribution since it represents the degree distribution of a RGG. This implies that  $\sigma_X = \bar{k}$ .

However, we can find the expression of distribution  $f_Y(y)$ , which is the distribution of neighbors degrees, using the relation:

$$f_Y(y) = \frac{f_Y(y|x=x^*)}{f_X(x|y=y^*)} f_X(x)$$

The numerator function  $f_Y(y|x=x^*)$  is the degree distribution of neighbors of a given node of degree  $x^*$ , and which has an expected value equal to  $k_{nn}(x^*)$ .

The other function  $f_X(x|y=y^*)$  represents the opposite case. This function is the probability distribution of nodes degrees given that they have a neighbor of degree equal to  $y^*$ .

Since the two functions are completely equivalent because of the symmetry of their definitions, we can conclude that  $f_Y(y) = f_X(x)$ , and, consequently,  $\sigma_X = \sigma_Y$ .

We can then conclude that  $r_2 \rightarrow A_2$  for large values of  $N$ .

In Fig. 4 we depict  $k_{nn}(k)$ , theoretical and empirical, in a RGG with  $N = 50000$  and  $\bar{k} = 50$ . From the figure, one can conclude that there is a very good agreement with the theoretical results.

This last result can be extended to other kinds of RGGs, with different dimensions or neighborhood volume (for  $d \geq 3$ ), since it does not depend on the shapes of the neighborhood volume, but only on the ratio of the average shared volume and the neighborhood volume. For the RGGs this ratio is intrinsically represented by the average clustering coefficient of the graph. The individual clustering coefficient  $C_X$  [1] of a node  $X$  is given by the following definition:

$$T_{XY} = \#\{\text{triangles with edge } (X, Y)\}$$

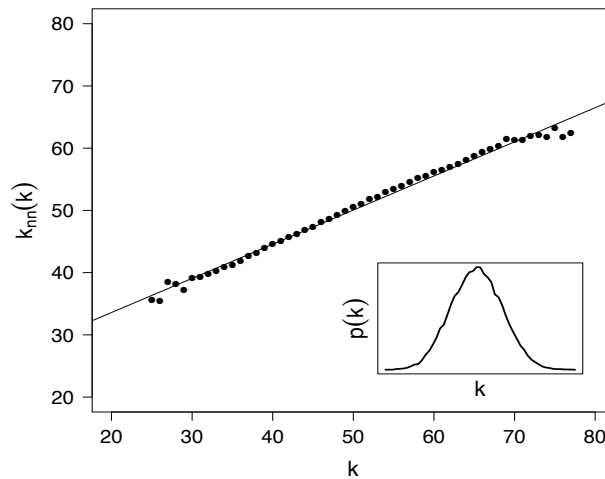


FIG. 4: Straight thick line: theoretical  $k_{nn}(k)$  with a slope coefficient of 0.586. Dotted curve: empirical  $k_{nn}(k)$ .  $N = 50000$ ,  $\bar{k} = 50$ . Due to the peaked empirical degree distribution function (see inset) low and high  $k$  values are noisier.

which represents the number of all triangles formed with the edge  $(X, Y)$  in the graph. It follows that:

$$C_X = \frac{1}{k_X(k_X - 1)} \sum_{Y \in V_X} T_{XY} \quad (1)$$

Let us denote  $A_d$  the ratio of the average shared volume of two connected nodes and the neighborhood volume of  $d$ -dimensional RGGs. We find that, on average,  $T_{XY} \rightarrow (k_X - 1)A_d$  and, substituting in equation (1),  $C_X \rightarrow A_d$ . Thus, the average clustering coefficient  $c_d = \frac{1}{N} \sum_X C_X$  of  $d$ -dimensional RGGs tends to  $A_d$ . Now, considering the formula which estimates  $c_d$  in the Euclidean space by Dall and Christensen [4]:

$$c_d \sim 3 \sqrt{\frac{2}{\pi d}} \left(\frac{3}{4}\right)^{\frac{d+1}{2}}$$

we can conclude that it represents a good approximation of  $A_d$  for large values of  $N$  and  $d$ , while, for small values of  $d$ , this function overestimates  $A_d$ .

The constant  $A_d$  depends on the neighborhood volume shape and represents the asymptotic value for the average clustering coefficient  $c_d$  and the assortativity coefficient  $r_d$ . The last assertion is due to the fact that, as we have seen above in the case  $d = 2$ ,  $r_2$  tends to the fraction  $A_2$ . This process is applicable to any dimension  $d$  in order to evaluate  $r_d$ .

Similar analytical results can be obtained in the same way extending to higher order of degree correlations, but the amount of calculus becomes particularly heavy. For example, the correlation coefficient between a given node's degree and the degree of its neighbors at distance 2 can be obtained from the study of the function  $k_{nn}^2(k)$ , which represents the average degree of neighbors at distance 2. Here the distance is intended to be the relational distance in the graph, i.e. the number of edges that compose the minimum shortest path which connects the two nodes. From Figs. 4 and 5 one sees that the degree correlations are non-negligible up to graph distance equal to 2 but they tend to disappear with increasing distance.

In summary, we have presented new results for the degree-degree correlations in RGGs in terms of their average clustering coefficients, showing exact results for the two-dimensional case and extending them to arbitrary finite dimension.

- 
- [1] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, UK, 2010.
  - [2] M. Barthélemy. Spatial networks. *Physics Reports*, 499:1–101, 2011.
  - [3] M. Penrose. *Random Geometric Graphs*. Oxford University Press, Oxford, UK, 2003.
  - [4] J. Dall and M. Christensen. Random geometric graphs. *Phys. Rev. E*, 66:016121, 2002.

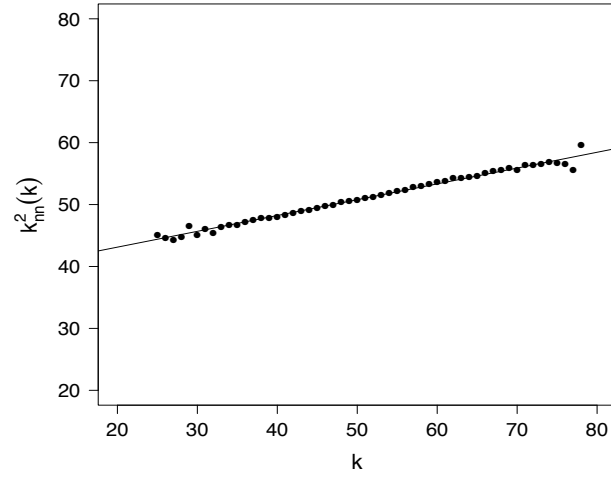


FIG. 5: Dotted curve: empirical  $k_{nn}^2(k)$ .  $N = 50000$ ,  $\bar{k} = 50$ . Straight thick line: linear regression line with slope coefficient of 0.255.

- [5] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.
- [6] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.*, 87:258701, 2001.